FDIC Banks Exploratory Data Analysis

Project Proposal

INTRODUCTION
The project (https://github.com/allaccountstaken/fdic_banks_eda) will provide explanatory data analysis (EDA) of banking industry historical evolution. It seems that the industry is going through a rapid consolidation as a number of regulated entities declined from 8,500 to roughly 5,000 in the last 20 years.

From a preliminary survey, it appears that only 500 banks actually failed during these years, and consequently, 3,500 disappeared due to other reasons, such as mergers, charter changes and voluntary liquidation. The main objective of the EDA is to create visual explanations of this consolidation, i.e. historical trends and drivers.

THE DATA
Federal Deposit Insurance Corporation (FDIC) is a United States government agency that examines and supervises the majority of commercial and savings banks (https://en.wikipedia.org/wiki/Federal_Deposit_Insurance_Corporation). The Corporation collects and maintains granular datasets describing the industry as a whole, as well as individual regulated entities.

Statutory Reports of Condition and Income (Call reports) are provided by the Examination Council via SOUP API: https://cdr.ffiec.gov/public/ManageFacsimiles.aspx. General data are available to the public via REST API: https://banks.data.fdic.gov/bankfind-suite/. Detailed Swagger documentation is available here: https://banks.data.fdic.gov/docs/swagger.yaml. Moreover, FDIC web sites contain numerous HTML tables with aggregated information that can be obtained programmatically with Python requests library.

Insights we are aiming to produce are important because credit institutions are critical for community development and economic growth. Moreover, costs of financial restructuring, although are not directly endured by tax payers, are still unproductive and avoidable. Finally, understanding consolidation drivers can help inform direct fintech innovation and future product development.

EXPERIMENTAL DESIGN
The data will be obtained from the official FDIC databases. Historical events will be queried using REST API and statutory reports will be obtained using SOAP connection. It is anticipated that two separated Python clients will be developed. Queries will be executed using report dates and unique entity identifiers, i.e. ID RSSD or FDIC Certificate Number.

It is also proposed that EDA will be following the waterfall chart pattern, https://en.wikipedia.org/wiki/Waterfall_chart, i.e. quarterly status: new entrants, mergers and spinoffs, and exits. This will require a tabular data format: rows being report dates and bank IDs, columns being event status and selected financial metrics. Expected format of the resulting table is provided below:

Report Date	Bank identifier (ID RSSD)	Events & Changes	Total Assets ($M)	Other financial metrics…
09/01/2019	12345XCV	New entrant	120	
01/01/2020	12345XCV	Active	150	
04/01/2020	12345XCV	Charter change	250	
07/01/2020	12345XCV	Liquidation	50	

PROJECT MANAGEMENT

Team member	Contacts	Task	Deliverables
Dima Mikhaylov	agp7dp	SOAP API client	Quarterly financial reports dataset
Cullan Bedwell	qja3fh	REST API client	Quarterly historical events dataset
Lauren Bassett	jme9rt	Pipeline integration	Data quality assurance
Casey Nguyen	ddr3de	Summary plots	Interactive plots, tentatively, a waterfall chart
Dima Mikhaylov	agp7dp	Command line interface	Manual querying of the dataset.
Cullan Bedwell	qja3fh	Data storage	Efficient data storage and retrieval tools

RESULTS

It is anticipated that the summary plot will allow for visual querying of the dataset, i.e. what entities were added and removed from the set during a specific time period. Alternatively, command line interface will allow for manual examination of the dataset.

TESTING

Task	Testing Strategy
SOAP API client	Unit tests, missing data check, data consistency checks
REST API client	Unit tests, missing data check, data consistency checks
Pipeline integration	Integration and regression testing, data quality checks, UAT
Summary plots	Dashboard design check, drill down, browser check, UAT
Command line interface	Unit tests, drill down, browser check, UAT
Data storage	Load and stress tests

OUTCOME
The final product will provide a unique big picture view of the banking industry, what will help general public better understand trends and drivers of the recent consolidation. Moreover, it is anticipated that users will enjoy extensive filtering functionality and easy to understand visuals. Finally, the resulting dataset can be used for multi-class classifications because it will contain labels describing historical events as well as relevant financial metrics.